Application of Fuzzy Differential Evolutionary Algorithms in Biological Data Mining

S.Poongothai[1,*], C.Dharuman[2], P. Venkatesan[3]

[1]Asst. Professor, Department of Mathematics, SRM University, Ramapuram Campus, Chennai, India.
[2] Professor, Department of Mathematics, SRM University, Ramapuram Campus, Chennai, India.
[3]Faculty of Research, Sri Ramachandra University, Chennai, India.
**Corresponding author: poongothaikannan25@gmail.com*

Abstract

In the world, everyday massive quantity of statistics being generated at a fee some distance higher than via which it could be tested by means of human comprehension alone, data mining became an essential undertaking for extracting as plenty useful data from this statistical data as possible. The standardized data mining techniques are appropriate enough to a positive extent but they are restrained by some limitations, and particularly for those cases that evolutionary approaches performs greater as well as extra capable. In this paper, a nature inspired evolutionary strategies is used for data mining with unique connection with Differential algorithm. We additionally expand to the hybrid algorithms- Fuzzy Differential Algorithms.

Keywords: *Data mining, Fuzzy logic, Differential Evolutionary algorithm, FURIA, Hybrid FDE*

1. Introduction

In data mining, especially in medical field, various computerized tools and techniques provides many advantages in health related problems. Most commonly used technique is artificial intelligence in disease diagnosis [1-3]. Evolutionary Algorithms are the class of general purpose of stochastic optimization algorithms based on neo-Darwinian theory. Differential Genetic Algorithm is one of the branch of Evolutionary Algorithm is recognized as a significant stochastic search technique in the recent world [4]. Here DE, which is more effective for feature selection or feature reduction that reduces the total number of attributes present in the dataset thereby reduces the time of searching and then fuzzy logic is applied for classification. The fuzzy logic is a tool used to find the optimal solution for the problems which deals with fuzzy input data [5]. Fuzzy logic is one of the best method of approaching the data mining process which involves computing the records primarily based at the likely predictions and clustering instead of normal conventional predictions. Fuzzy logic based Algorithms are increasingly being carried out and applied in several disciplines to help in datamining [6].These FL based algorithms are widely used in medical field to predict the disease earlier so that the specialists enable to detect and evaluate the risks of the disease. On the other hand, early detection of the disease is one of the vital approach of curing the ailment.

1.1 Motivational database

The dataset of Lung cancer is considered for this paper that has taken from UCI Machine Learning Repository . It consists of fifty seven attributes including class lable with thirty two instances. Recently, Lung cancer is considered as one of the major health problems which affects the human globally [7]. Eventhough there are lot of techniques available in the early detection of disease, the huge amount of data makes difficult to evaluate it [8]. Feature reduction or feature selection makes the evaluation of data easier and also cost effective.

2. Preliminaries

2.1 Fuzzy Logic

Fuzzy set theory shows extremely powerful for solving the uncertainty in the problem. It is defined for "vagueness". It is an outstanding mathematical method which shows how to handle the uncertainty arising due to vagueness. Mathematically, a fuzzy set A is defined to be a set of all ordered pairs,

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

where X is the universe of discourse and $\mu_A(x)$ is called the membership function of respective x in A. For each element of X, the membership function assigned to the element between zero and one. In general, Fuzzy logic is considered as a set of mathematical rules and ideas based on degrees of membership functions for expertise the illustrations [10]. Fuzzy rule based systems are actually the extension of classical rule based system. Fuzzy logic plays as an effective predictor of the disease. The hybridization of Fuzzy Logic and Evolutionary Algorithms is the most popular approach in the research world which leads to Genetic Fuzzy Systems (GFSs) [11] and Fuzzy EAs [12, 13, 14]. Fuzzy sets and fuzzy logic performs better to classify the data. compared to crisp classifications, fuzzy classification are easily allows for analyzing the data samples in a better way. Fuzzy inference system is an example of efficient, excellent and easily interpretable fuzzy classifiers [15,16]. Since IF-THEN rules are comprehensible to a human being, it is considered as the most popular way of expressing the knowledge in the field of machine learning and data mining [17]. The fuzzy inference system (FIS) follows the technique as "If antecedent then consequent" [18,19].

2.1 Differential Evolutionary Algorithms

Evolutionary Algorithms for solving NP-complete problems have become a more popular research area in these days. One of the branch of Evolutionary Algorithms is Differential Evolution (DE) which was developed by Rainer Storn and Kenneth Price [9] to optimize the problems. The main advantages of DE are its, easily applicable, fast and accuracy. DE is one of the high-quality evolutionary algorithms for solving real life problems. DE has been utilized in numerous technology and various applications in engineering fields to find out powerful answers even for complex problems. If a gadget is amenable to being rationally evaluated, DE can provide the method for extracting the quality viable overall performance from it .Mutation is considered as a main operator in DE for searching and selection is the operator to find the solution in the feasible region. DE Algorithms will generate randomly a sequence of populations by using *selection mechanisms* and *crossover* and *mutation* are applied as search mechanisms. The flowchart of DE is shown in figure 1.

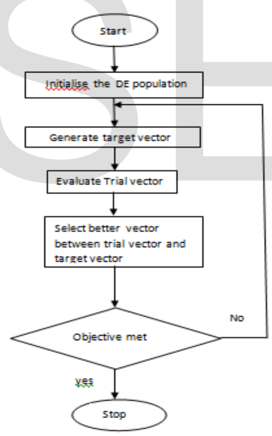

Fig. 1. Flowchart of DE.

3. Methods and Results

Consider a population size of *N*. The population matrix can be taken as $y^g_{m,i} = [y^g_{m,1}, y^g_{m,2}, y^g_{m,3}, \cdots y^g_{m,n}]$ where, *g* is the Generation and m=1,2,3,…*N*. Initial population is generated randomly between upper and lower bound. Let the initial population be generated as

$$y_{m,i} = y^L_{m,i} + RAND\left(y^U_{m,i} - y^L_{m,i}\right), i = 1,2,\dots,n \ \ and \ \ m = 1,2,\dots,N \qquad (1)$$

where $y_{m,i}^{U}$ is the upper bound of the variable y_i and $y_{m,i}^{L}$ is the lower bound of the variable y_i. The differential mutation operator generates the mutated variable $\mu_{m,i}^{g+1}$, by the equation:

$$\mu_{m,i}^{g+1} = y_{R1,g} + S\left(y_{R2,g} - y_{R3,g}\right)$$

where R_1, R_2, R_3 are not equal to each other and S is a real number in the interval [0,2]. DE then applies a crossover operation where the mutated vector is combined with the principal vector (parent) and produces the new vector known as trial vector and is defined as

$$\tau_i^{g+1} = \begin{Bmatrix} \mu_{i,t}^{g+1} \text{ } if \text{ } RAND\,(t) < CR \text{ } or \text{ } t = Rm_i \\ y_{i,t}^{g} \text{ } if \text{ } RAND\,(t) > CR \text{ } and \text{ } t \neq Rm_i \end{Bmatrix}$$

where CR is a crossover constant belongs to the interval [0,1], t=1,2,..., n and $RAND(t)$ is the t^th^ evaluation of the generator number in the same interval. At last, a index which is randomly taken $Rm_i \in [1,2,\ldots,n]$ is used to ensure that τ_i^{g+1} has atleast one element from μ_i^{g+1}. Then selection operator is adapted by DE to confirm that the vectors get improved to the next level iteration or generation and they are definitely better ones in the population compared to the previous generation. The iterations are repeated for some number of generations till the objective condition satisfied. By this method the attributes are reduced to 20. Figure 2 shows the classification rate of Fuzzy.

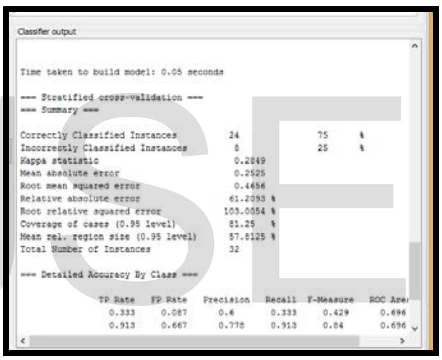

Fig. 2. Classification rate of Fuzzy Rules.

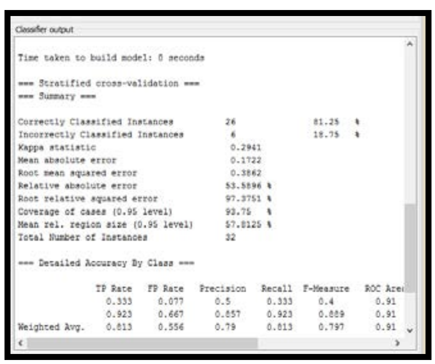

Fig. 3. Classification rate of FDE.

The reduced variables or attributes plays an input for Fuzzy systems. In this paper, *Fuzzy Unordered Rule Induction Algorithm* (FURIA) i.e., fuzzy rule-based classification method is used [20]. It is an extension of Ripper Algorithm which used to study fuzzy rules instead of conventional by applying the efficient rule stretching method. The classification results obtained by using data sets that are frequently employed in the corresponding literature suggest that the FURIA algorithm is competitive and has good classification accuracy. To evaluate classification accuracy, it is crucial to consider the number of used attributes and the number of induced rules that classify the data set. It can be considered that the rule-stretching feature of FURIA could provide a significant advantage in data sets where there would be a need to classify records that differ from those used in training. Figure 3 shows the classification rate of Fuzzy DE. Comparison of Fuzzy and FDE obtained by 10-fold cross validation is shown in Table 1.

Table 1. Comparison of Fuzzy and FDE.

Algorithms	Classification rate (%)	Mean Absolute Error
Fuzzy	75.00	0.2525
Fuzzy DE	81.25	0.1722

4. Discussions and Conclusions

From Table 1 we note that the Fuzzy DE performs better than Fuzzy classification rule. The classification is done by FURIA and DE reduced 57 attributes to 20 attributes and the results are compared with and without hybridization which shows that the idea of hybridization of DE with fuzzy sets in a promising way. The classification rate of Fuzzy and FDE are 75% and 81.25% respectively. Hence hybridization yields better classification compared to other normal methods. In future, Type-2 Fuzzy sets can be considered for hybridization with other types of Evolutionary Algorithms.

References

[1] Overiu M., Simon D.,2010, " Biogeography-based optimization of neuro-fuzzy system arameters for diagnosis of Cardiac disease", Proceedings of the 12th annual conference on Genetic and evolutionary computation, p. 1235–42.
[2] Rajeswari K., Vaithiyanathan V., Amirtharaj P., 2011, " Prediction of risk score for heart disease in India using Machine intelligence", International Conference on Information and Network Technology; Singapore: IACSIT Press; p.19–22.
[3] Srimani P.K., Koti M.S.,2014, " Knowledge discovery in medical data by using rough set rule induction algorithms", Indian Journal of Science and Technology, Jul; 7(7):905–15.
[4] Fogel, D. B. (2006). Evolutionary computation: toward a new philosophy of machine intelligence, John Wiley and Sons.
[5] Bojadziev, G. and M. Bojadziev (1995). Fuzzy Sets, Fuzzy Logic, Applications. Singapore, World Scientific.
[6] Zadeh, L. A. (1965). "Fuzzy sets." Information and control 8: 338-358.
[7] Jinyan Li, Huiqing Liu, James R Downing, Allen Eng-Juh Yeoh, Limsoon Wong (2003). Simple rules underlying gene expres sion profiles of more than six subtypes of acute lymphoblastic leukemia (ALL) patients Bioinformatics vol 19, pages 71 2003.
[8] Saleem Durai, M.A. and Iyengar, N.Ch.S.N., "Effective Analysis and Diagnosis of Lung Cancer Using Fuzzy Rules" International Journal of Engineering Science and Technology, Vol. 2 (2010), Page no. 2102-2108.
[9] Price, K., R. Storn, et al. (2005). Differential evolution: A practical approach to global optimization. New York, Springer. Price and Storn, 1997.
[10] Hong T., 2008, " MCMC algorithm, integrated four-dimensional seismic reservoir characterization and uncertainty analysis in a bayesian framework", ProQuest LLC, p. 31.
[11] O. Cordon, F. Herrera, F. Hoffmann and L.Magdalena, Genetic fuzzy systems. In: Evolutionary tuning and learning of fuzzy knowledge bases, World Scientific, Singapore (2001).
[12] F. Herrera, M. Lozano and J.L. Verdegay, Tackling fuzzy genetic algorithms. In: Genetic Algorithms in Engineering and Computer Science,JohnWiley, New York. (1995), 167-189.
[13] W. Pedrycz, Fuzzy evolutionary computing. Soft Computing 2, (1998), 61-72.
[14] A. Tettamanziand M. Tomassini, Fuzzy evolutionary algorithms. In: Soft Computing: Integrating Evolutionary, Neural, and Fuzzy Systems, Springer, Heidelberg. (2001), 233- 248.
[15] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. R. Pal, Fuzzy Models and Algorithms for Pattern Recognition and Image Processing (TheHandbooks of Fuzzy Sets). Secaucus, NJ, USA: Springer-Verlag NewYork, Inc., 2005.
[16] A. Verikas, J. Guzaitis, A. Gelzinis, and M. Bacauskiene, "A general framework for designing a fuzzy rule-based classifier," Knowledge and Information Systems, (2010), 1-19.
[17] U.M. Fayyad, G. Piatetsky-Shapiro and P. Smyth. From data mining to knowledge discovery: an overview. In: U.M. Fayyad et al. (Eds.) Advances in Knowledge Discovery and Data Mining, AAAI/MIT, (1996), 1-34.
[18] T.Hong,MCMC algorithm, integrated four-dimensional seismic reservoir characterization and uncertainty analysis in a Bayesian framework, ProQuest LLC, (2008) p. 31.
[19] S.Poongothai, C.Dharuman and P.Venkatesan, Fuzzy Evolutionary Computing in Biological Data Mining, Global Journal of Pure and Applied Mathematics (GJPAM) ISSN 0973-1768, 12, No. 1(2016), 306-309.
[20] Cohen W. W., Fast Effective Rule Induction. *Proceedings of the Twelfth International Conference on Machine Learning*, 1995, pp. 115-123.